

AI-Enabled Mission Engineering

Michael Pennock, Jesse Ponnock, Trac Bannon, Katja Sednew, Patty McDermott, Judith Dahmann

The MITRE Corporation

AI4SE & SE4AI Research Application Workshop 2026

September 22-23, 2026

Approved for Public Release; Distribution Unlimited. Public Release Case Number 26-1832

NOTICE

Portions of this technical data were produced for the U.S. Government under Contract No. FA8702-19-C-0001 and W56KGU-18-D-0004, and is subject to the Rights in Technical Data-Noncommercial Items Clause DFARS 252.227-7013 (FEB 2014)

The work was supported by the MITRE Independent Research and Development Program

© 2026 The MITRE Corporation

Project Team

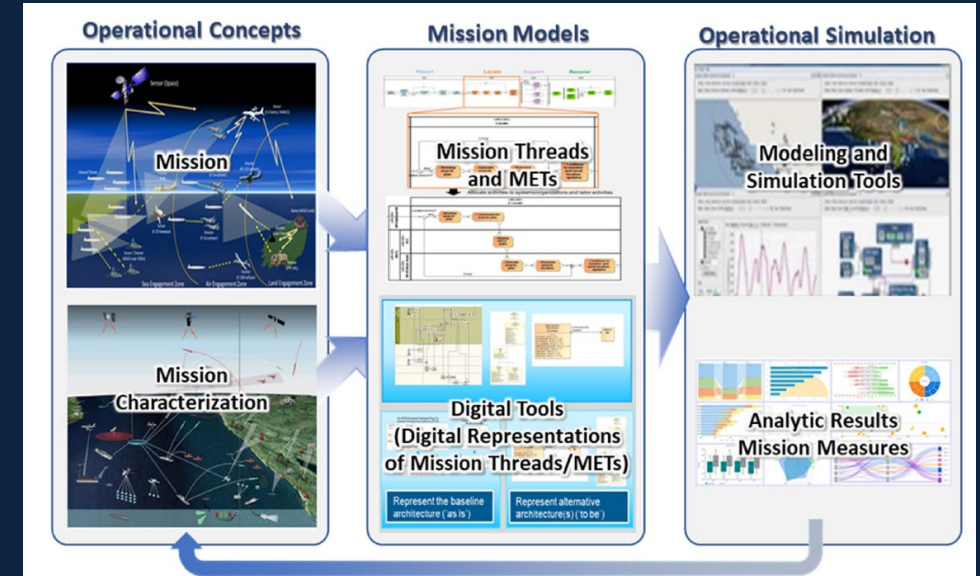
Intentionally interdisciplinary team of engineers and scientists: Data Scientists, Mission, Simulation, Human Factors, Systems, and Software Engineers

- Aanchal Mathur
- Alex Cooper
- Adam Barnes
- Amy Marsh
- Andrew Sillers
- Arees Agobian
- Benjamin Yam
- Bridget Musselman
- Christina Thompson
- Dominick Taylor
- Don Kim
- Drew Founds
- Faith Morgan
- Gabriela Parasidis
- Jack Thompson
- Javier Aleman
- Jerry Schweiger
- Jesse Ponnock
- Joshua Lee
- Judith Dahmann
- Justin Lieffers
- Kat Mangan
- Katja Sednew
- Huma Chowdhury
- Leah Balkin
- Marcel Marki
- Mia Barger
- Michael Briggs
- Michael Pennock
- Michael Perez
- Nicholas Kinberg
- Patty McDermott
- Pete Korfiatis
- Rob Scheible
- Saina Shibili
- Sean Williams
- Son Huynh
- Steele Farnsworth
- Trac Bannon
- William Macke
- Yakov Neshcheretnyy

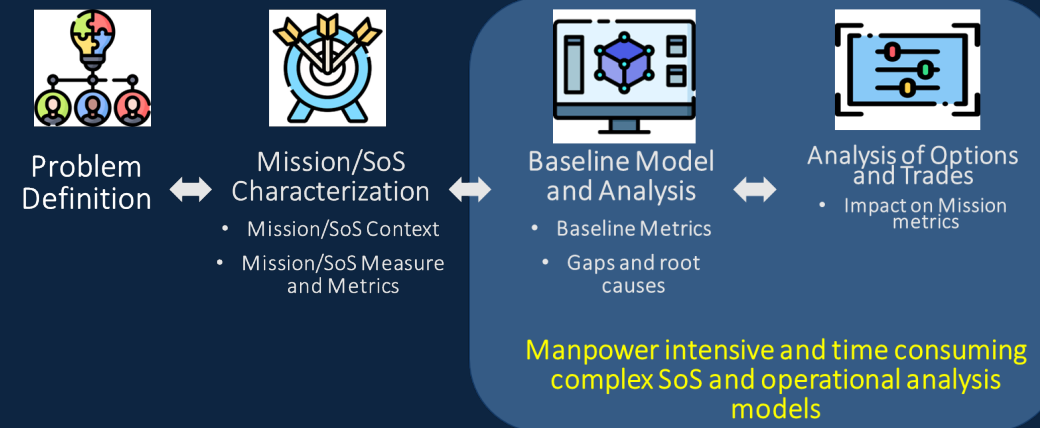
Government Sponsor Advocate: Marc Goldenberg, Principal Deputy, Mission Engineering and Integration Activity

Project Summary

- Mission engineering (ME) is essential for the US DoW to make the **right investments to ensure warfighter success**
- Today, the efficacy of mission engineering (ME) is limited by our ability to understand **complex mission architectures**
- To address this complexity, our project is teaching LLMs how to facilitate **conversations with a digital mission model**
- We are **adapting commercial AI capabilities** and evaluating their impact on ME workflows using a combination of **structured user testing** and application to **real mission engineering projects**



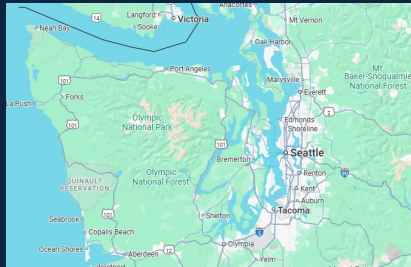
Source: DoD MEG 2.0, 2023



DoW's ME Challenge

Digital engineering models are central to the mission engineering workflow

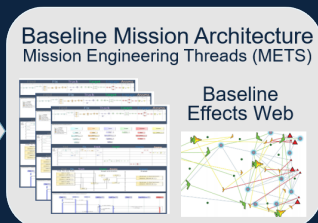
Defense of Seattle*



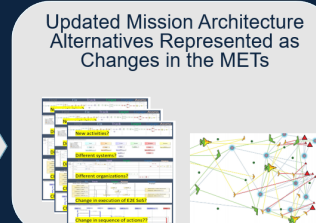
Obtain Mission Thread Source Information (i.e., OPLAN) For selected scenario/vignette



Informs



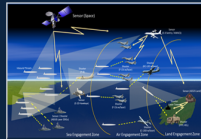
Provides basis for



Represented in AFSIM

Represented in AFSIM

Alignment with Scenario Documentation (i.e., JFOS)



Informs

Analysis of Baseline Compared to Alternative on Mission Outcome Metrics In Selected Scenario



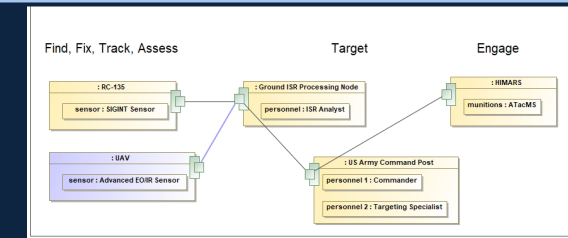
Operational Mission Analysis



Baseline

Alternative

Model the SoS in an architecture tool

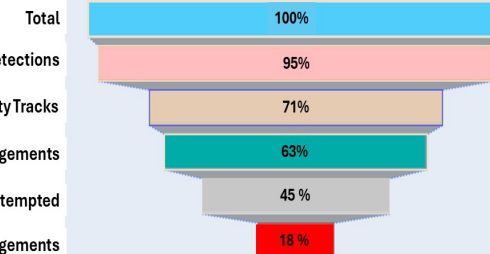


Model and analyze the relevant portions of the SoS using an operational simulation tool

* Source: Dahmann, D. J. S., & Parasidis, G. I. (2024). Mission Engineering. *The ITEA Journal of Test and Evaluation*, 45(3).

Our success at digital engineering has created a new challenge: large, complex digital models are difficult to review and extract valuable insights

Adversary Targets



```
1 class CodeExample():
2     def printStatement(self):
3         print('Hello World!')
4
5 def main():
6     classEx = CodeExample()
7     classEx.printStatement()
8     if __name__ == "__main__":
9         main()
```

[File:Hello World in Python.png - Wikimedia Commons](#)

MITRE AI4ME Research Addresses Model Complexity

	What are the contents of a model?	What elements need to be adapted or restructured?	Is the model consistent with key artifacts?
Mission Engineer	Assess progress and completeness before delivery Extract data from models to use in a new model	Understand what needs to change to architect and analyze alternatives to baseline Identify gaps in externally provided models	Check architecture and simulation models for mutual consistency Check model for performance aligned to system test data
Prospective Model User	Learn organization and composition of models	Learn what parts of the model align with new use	Check model against new user considerations
Reviewer/ Sponsor	Verify that model contains required components	Recommend model improvements	Verify model is consistent with authoritative sources

Employ commercial AI/LLM within defense mission context

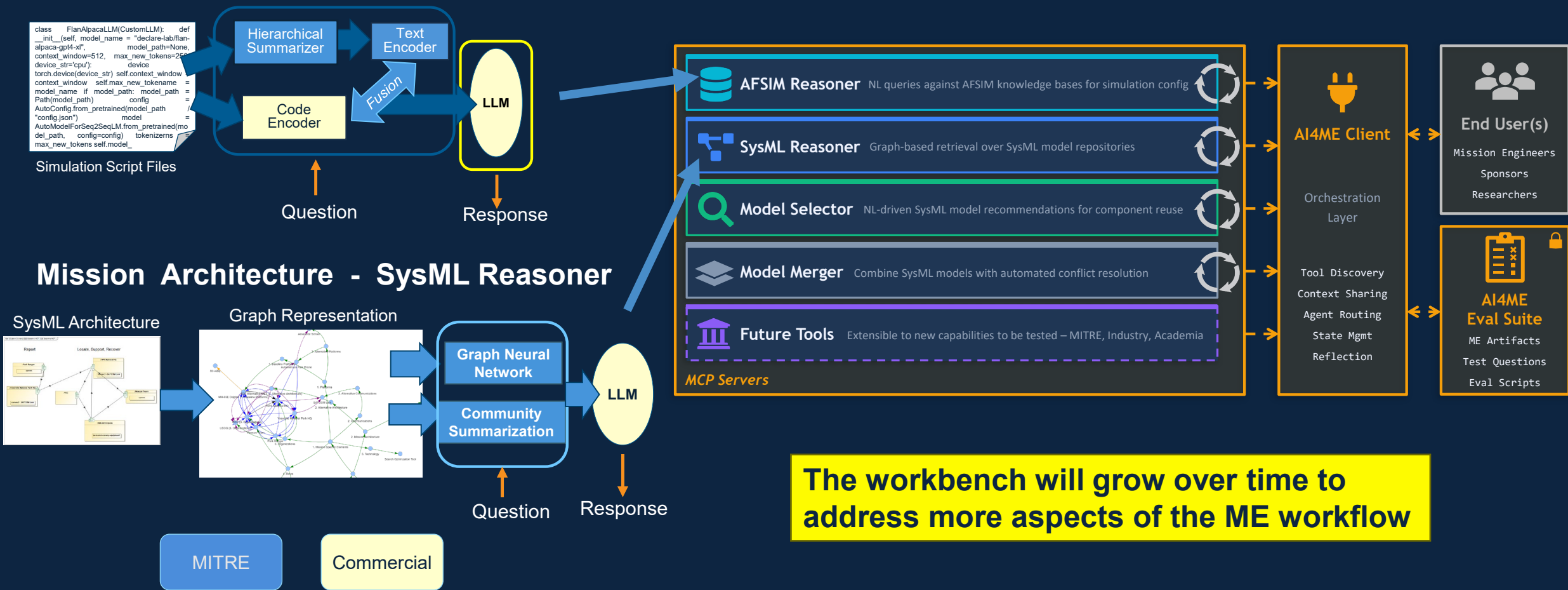
Address ME model complexity

- Transparency
- Accessibility
- Extensibility
- Reusability
- Trustworthiness

Focus on mission architectures in SysML and operational simulation script files

Building out an AI4ME Workbench

Operational Simulation Reasoner



The workbench will grow over time to address more aspects of the ME workflow

* Note: AFSIM is a commonly used operational simulation in the DoW

Applying Tools to Mission Engineering Projects

AFSIM Reasoner Tool

what are the contents of this software module

Based on the context provided, the software module in question appears to be part of the Joplin project within the AFSIM framework. Here's a detailed breakdown of the contents and functionalities of this module:

Overview

The Joplin project is designed to enhance simulation environments by integrating advanced natural language processing (NLP) and retrieval-augmented generation (RAG) techniques. It supports complex text processing tasks, enabling simulations to effectively process and understand natural language inputs. The modular design allows for future extensions, making it a flexible and scalable component within the AFSIM environment.

Key Components

- Evaluation and Benchmarking:**
 - benchmark.py:** Evaluates and benchmarks simulations using various evaluators and metrics to assess language model outputs, focusing on relevance, correctness, and faithfulness.
- Prompt Templates:**
 - prompt_templates.py:** Dynamically selects prompt templates based on the model name, optimizing prompts for different language models like LLAMA3, MIXTRAL, and GPT.
- Language Model Integration:**
 - llm.py:** Integrates a language model to generate text completions, enhancing text-based interactions within

* Note: example uses surrogate Python code

SysML Reasoner Tool

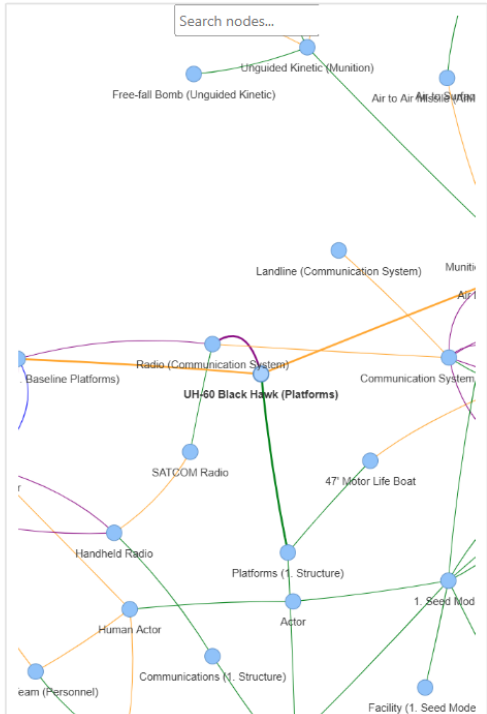
Select ME Model
Stranded Hiker

Query about the graph
Is there a UH-60 in the mission architecture?

Submit Query

Detected Entities
["UH-60 Black Hawk (Platforms)", "2. Mission Architecture"]

Query Results
Yes, the UH-60 Black Hawk is present in the mission architecture. It is classified as an airborne platform and is used in various operations, including those involving rescue missions and coordination with the U.S. Coast Guard. The Black Hawk helicopter is outfitted with a communication system that facilitates radio and satellite communications, which are critical for coordinating with rescue teams and other organizations like the National Park Service and the Rescue Coordination Center. Furthermore, the UH-60 Black Hawk supports operations by transporting rescue teams, such as those from the Park Rangers, to necessary

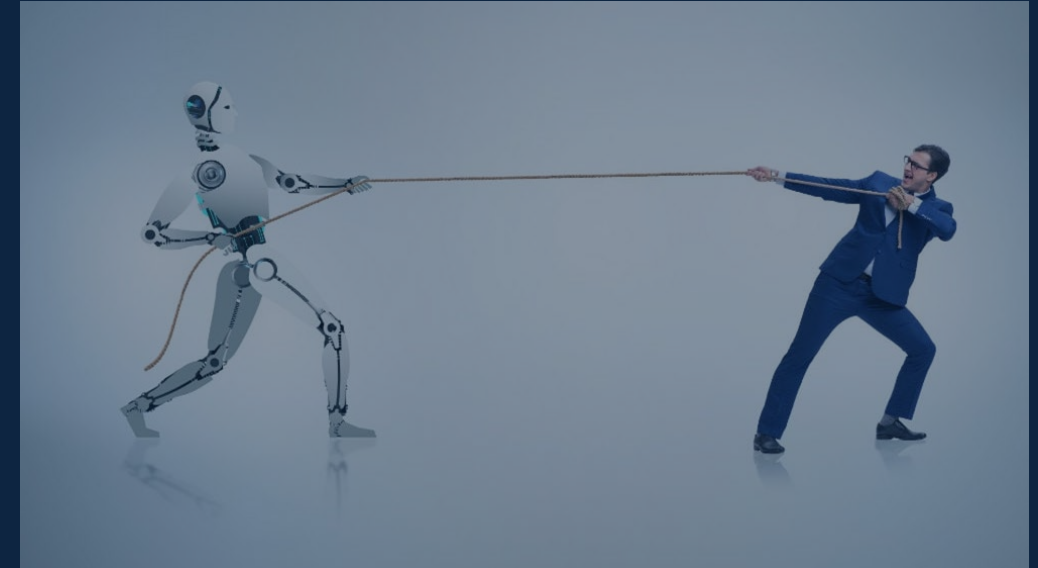


The SysML diagram illustrates a mission architecture with various components and their interconnections. Key components include: 'Free-fall Bomb (Unguided Kinetic)', 'Unguided Kinetic (Munition)', 'Air to Air Munition', 'Landline (Communication System)', 'Munition', 'Baseline Platforms', 'Radio (Communication System)', 'UH-60 Black Hawk (Platforms)', 'Communication System', 'SATCOM Radio', '47' Motor Life Boat', 'Handheld Radio', 'Platforms (1. Structure)', 'Actor', 'Human Actor', 'Communications (1. Structure)', 'Team (Personnel)', '1. Seed Mod', and 'Facility (1. Seed Mod)'. The diagram uses colored lines to represent different types of relationships or data flows between these entities.

* Note: example uses a surrogate but realistic ME model

SysML User Study

- **Purpose:** Examine whether access to an AI tool changes how mission engineers perform a new modeling task.
- **Research question:** How does an AI-enabled workflow affect task completion, understanding, and confidence?
- **Approach:**
 - Mission engineers completed two similar tasks
 - Each task included a baseline OV-1, an alternative OV-1, and previous models available for reuse
 - Each participant completed one task with access to an AI tool and one task without it
 - We randomized both task order and which task was paired with AI access
 - Each task lasted ~3 hours, followed by questionnaires



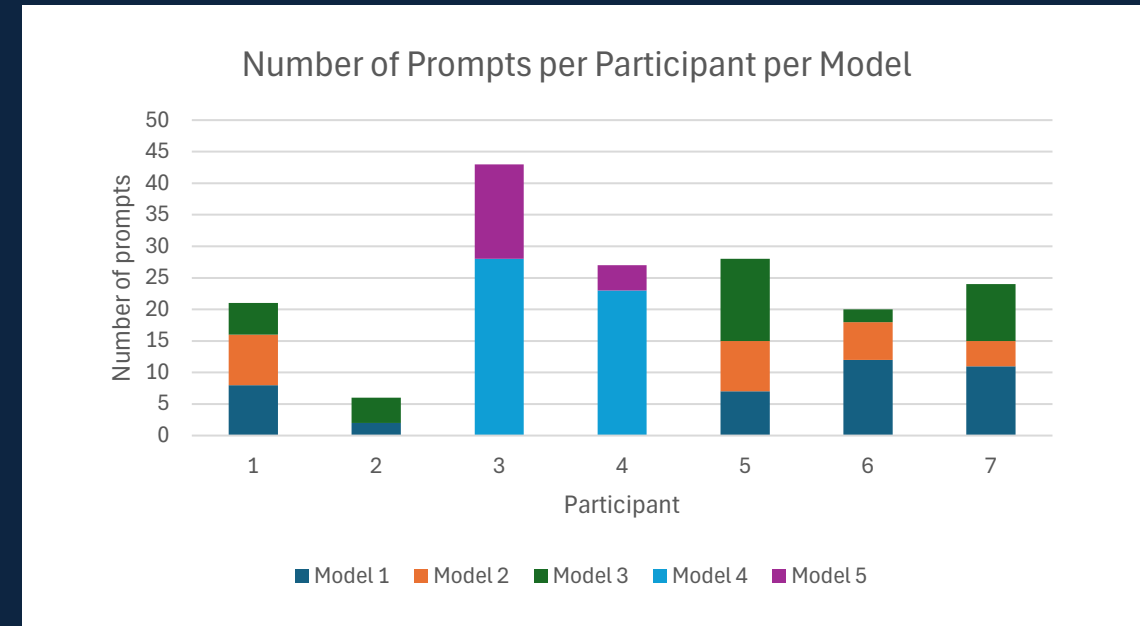
SysML user study examines uplift from AI for mission modeling. It is not AI vs. humans.

User Study: Mission Engineer Use of SysML Reasoner

AI-Enabled Workflow: Type of Queries

- Understand connections or interdependencies
 - For the baseline scenario, identify systems connected to the F-15.
 - How do naval fighters interact with the CVN?
- Understand the scenario or mission
 - What is the description of the [xyz] mission?
 - Where does AWACS deploy from?
- Reusability
 - Which elements would be candidates for reuse as a C2 center?
- Understand model contents at a high level
 - Please give me a list of the systems employed in the baseline mission engineering thread.
- Pinpoint a specific entity in a model
 - What are the names of naval fighters?
- Identify sub-systems or sub-components
 - What are the parts of the F-18 air vehicle system?
 - Does the F-18 have multi-channel comms?

AI-Enabled Workflow: Frequency and timing of tool use across the session*

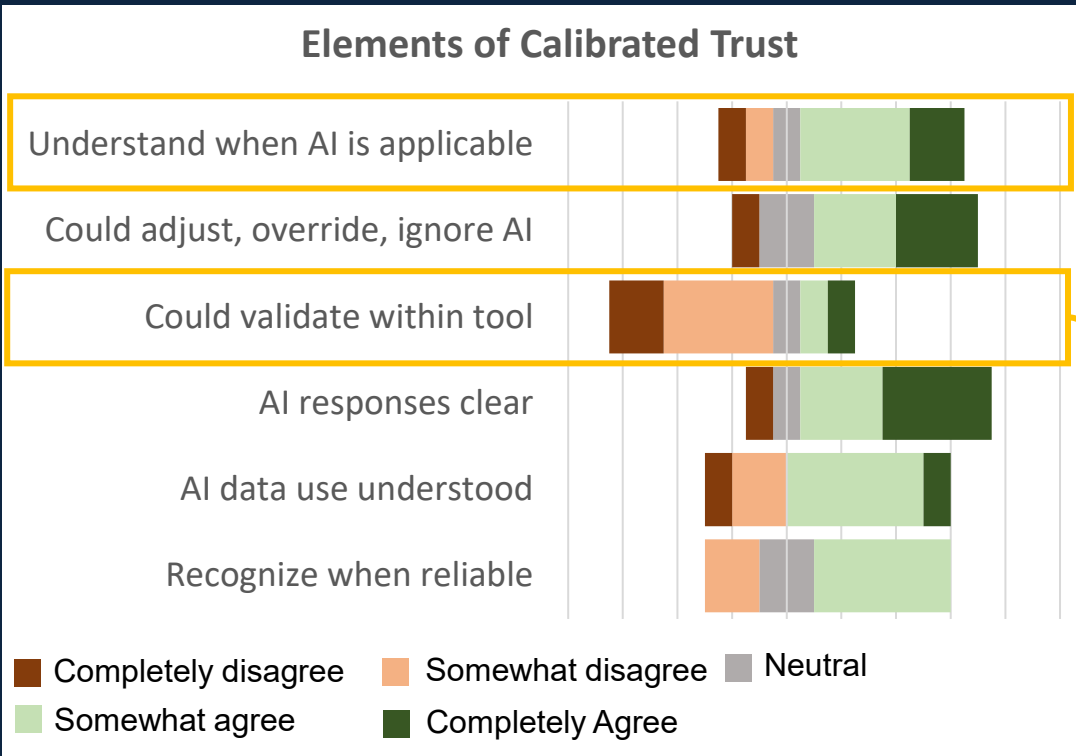


*Note that two participants' data is missing due to a logging error.

User Study: Calibrated Trust

Moderately High Overall Trust (5.5 on a 7-point scale where 7 = a great deal of trust in the system)

Elements of Calibrated Trust



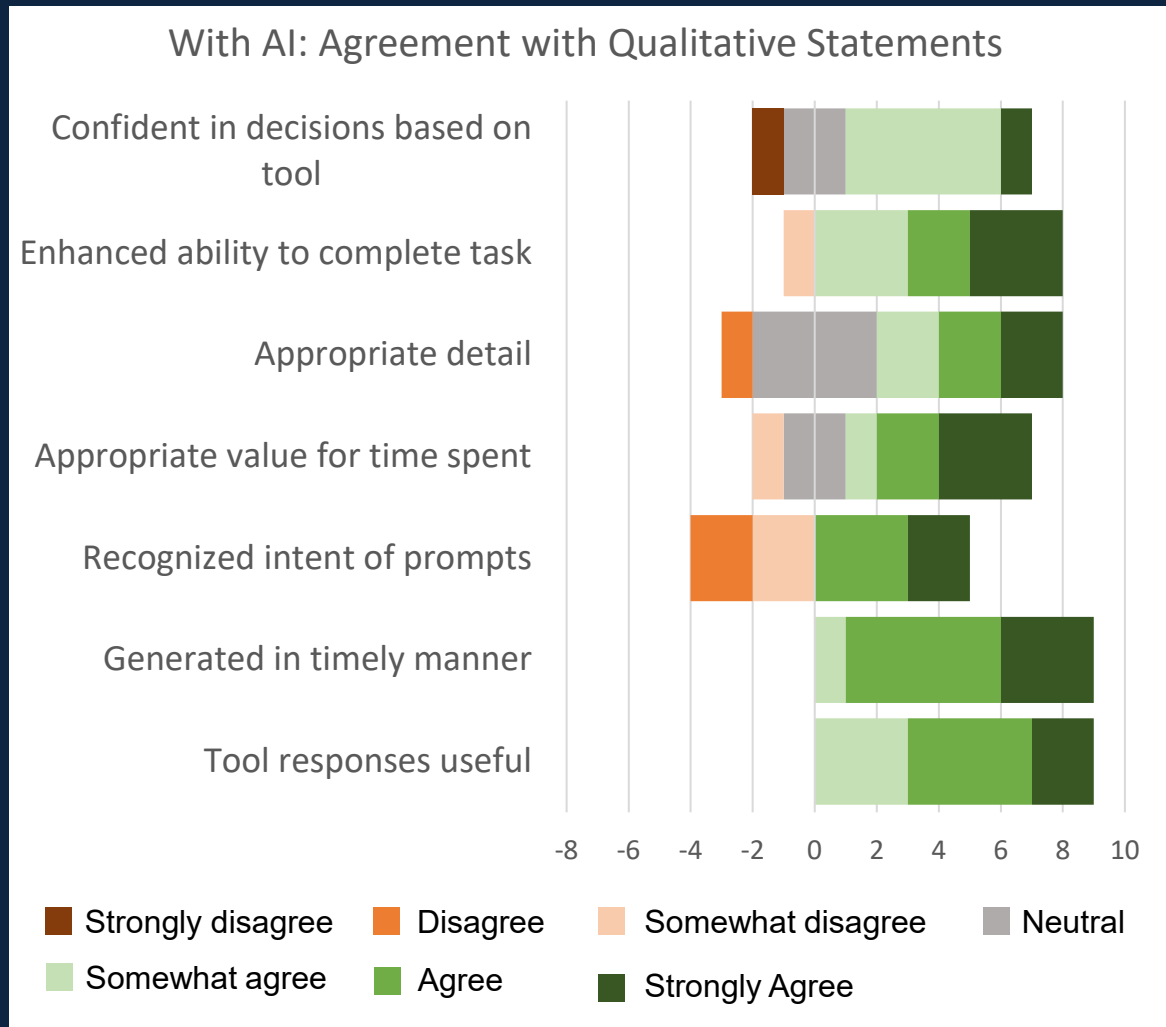
Support for Calibrated Trust

- **6/9 mission engineers understood situations in which the SysML Reasoner is applicable**
 - “It is a good search tool” “it is good at questions about model specifics”
 - “It is not as good at generic questions” “It wasn’t helpful with questions about system relationships”
- **Tool was Directable (could ignore, adjust)**
- **AI responses were clear and explainable**
- **AI data use was clear**

Opportunity for Improvement

- **Participants could not validate AI responses within the tool, *but they did use the original models to check the AI responses***

User Study: Perceptions of AI Impact on Workflow



Mission Engineers believed the SysML Reasoner

- Enhanced ability to complete the task
- Provided value appropriate for the time spent
- Generated responses in a timely manner
- Provided useful responses

Opportunity for Improvement

- Half the mission engineers wanted more detailed responses
- 4/9 mission engineers found that the SysML Reasoner did not recognize the intent of their prompts

We are looking to collaborate

- Goal: Learn how real mission engineers can benefit from AI solutions
 - Do you have ME workflow challenges that you recommend we investigate?
 - Do you have ME efforts that might benefit from experimental use of AI?
- Goal: Expand the workbench with additional capabilities
 - Do you have AI capabilities you think could benefit ME but no ME data to test with?
- Goal: Develop approaches for AI-assisted mission model construction
 - Do you have real examples of ME digital model integration challenges that could be used to train and test AI capabilities?

**We are looking for government and industry partners to
grow the AI4ME Workbench**

Michael Pennock
Mission Engineering Capability Chief Engineer
The MITRE Corporation
mpennock@mitre.org

Backup

Research Questions

- CY25 research question:
 - Can fine-tuning and/or augmentation of selected LLMs enable mission engineers to accurately search and summarize technical artifacts?
- CY26 research question:
 - Can emerging AI technologies enable MEs to identify and assemble existing mission model components for a new mission problem?

SysML User Study

- Completed initial analysis:
 - Workflow patterns when mission engineers had access to the SysML Reasoner
 - Perceptions of usefulness, confidence, and impact of the SysML Reasoner
- The full analysis will expand this to examine the impact of having access to the SysML Reasoner on
 - Model understanding
 - Completeness/quality of the results
 - Task time
- The end goal is to compare AI-enabled and non-AI workflows in terms of both process and outcomes